Quality of Spirometry Test Performance in Children and Adolescents*

Experience in a Large Field Study

Paul L. Enright, MD; William S. Linn, MA; Edward L. Avol, MS; Helene G. Margolis, MA; Henry Gong, Jr, MD, FCCP; and John M. Peters, MD, ScD

Study objective: **To determine the ability of children and adolescents to meet the American Thoracic Society (ATS) goals for spirometry quality that were based on results from adults.**
Design: **Observational.**
Participants: **More than 4,000 public school students, ages 9 to 18 years.**
Measurements: **Spirometry was performed annually for 3 years, with the recording of maneuver quality measures of forced expiratory time, end-of-test volume, back-extrapolated volume, and time to peak expiratory flow (PEFT), and the recording of differences between best and second-best FVC, FEV_1, and peak expiratory flow (PEF) values.**
Results: **Regression analyses showed significant influences of participant age, gender, ethnicity, size, clinical status, and previous testing experience, as well as differences among individual test technicians. In general, these influences were small and explained little of the variance in performance. On average, children with a history of asthma or wheeze performed better quality spirometry than did others. Only PEFT improved significantly from year to year. Overall, only 15% of girls' tests and 32% of boys' tests met the PEFT criterion derived from adults in the Lung Health Study.**
Conclusion: **Most of the children met adult-based ATS goals for spirometry test performance. Age group-specific criteria are needed to ensure adequately fast PEFT and reproducible PEF values.**
(CHEST 2000; 118:665–671)

Key words: children; peak flow; quality control; spirometry

Abbreviations: ATS = American Thoracic Society; BEV = back extrapolated volume; BMI = body mass index (kilograms/meters squared); $dFEV_1$ = the difference between the best and second-best FEV_1 value within a spirometry test session; dFVC = the difference between the best and second-best FVC value within a spirometry test session; dPEF = the difference between the best and second-best peak expiratory flow value within a spirometry test session; EOTV = end-of-test volume; FET = forced expiratory time; LHS = Lung Health Study; PEF = peak expiratory flow; PEFT = time to peak expiratory flow; QA = quality assurance

In epidemiologic studies in which spirometry results are a primary outcome measurement, the results depend not only on the true lung function of the subject populations, but also on the quality of their test performance. Accordingly, a strong quality assurance (QA) program is essential. Poor test session quality should be documented, because it may tend to obscure the relationship between risk factors and respiratory health or may itself be an index of impaired respiratory health.

*From the University of Arizona (Dr. Enright), Tucson, AZ; Department of Preventive Medicine (Messrs. Linn and Avol, and Drs. Gong and Peters), University of Southern California, Los Angeles; and California Air Resources Board (Ms. Margolis), Sacramento, CA.
Supported by the California Air Resources Board (contract No. A033–186), the National Institute of Environmental Health Sciences (grant No. 5P30ES07048–02), the US Environmental Protection Agency (contract No. CR824034–01-3), and the Hastings Foundation.

The statements and conclusions in this report are those of the authors and not necessarily those of the California Air Resources Board (or other sponsors). The mention of commercial products and their source or use in connection with the material reported herein is not to be construed as actual or implied endorsement of such products.
Manuscript received September 28, 1999; revision accepted April 19, 2000.
Correspondence to: Paul Enright, MD, University of Arizona, HSC 2342, 1501 N Campbell Ave, Tucson, AZ 85724; e-mail: lungguy@aol.com

Downloaded From: http://journal.publications.chestnet.org/pdfaccess.ashx?url=/data/journals/chest/21952/ on 01/05/2017

The 1994 American Thoracic Society (ATS) criteria for satisfactory spirometry performance[1] are based almost exclusively on studies of adults. To determine whether these criteria (goals) are suitable for younger populations, we reviewed the spirometry data from 3 successive years of testing in the Children's Health Study,[2,3] a 10-year longitudinal investigation of air pollution and other environmental risk factors, which involves several thousand subjects aged 9 to 18 years.

Materials and Methods

Students from public schools in 12 middle-income communities with diverse air pollution patterns in southern California were studied. In year 1 (1993), students in grades 4, 7, and 10 were enrolled. In year 4, a new cohort of grade 4 students was enrolled. Subjects were evaluated annually from January through June by spirometry and by questionnaires on respiratory health and environmental exposures. They were tested at their schools, usually in groups of ≥ 12. As many as six technicians performed spirometric tests concurrently in the same room, each using a dry rolling seal spirometer with a personal computer-based data acquisition system (Spiroflow model 132; PK Morgan Ltd; Gillingham, UK). Coaching, data acquisition, and data management techniques were similar to those in the Six Cities Study.[4] Frequent checks against a 3.00-L calibration syringe (Flow Volume Calibrator; Jones Medical Instrument; Oak Brook, IL) ensured that data were consistent between different spirometers and across time.[5] In year 6 only, about 4% of tests were performed at the homes of subjects who had moved away from the study school districts.

In each annual test session, the subject was vigorously coached by the technician to perform forced expirations until three acceptable maneuvers (or a maximum of seven) were recorded. The initial criteria for an acceptable maneuver were the following: (1) FVC and FEV_1 values that were at least 95% of the largest values; (2) back-extrapolation volume (BEV) < 5% of FVC; (3) end-of-test volume (EOTV) of < 50 mL in the final 2 s; and (4) forced expiratory time (FET) of at least 6 s. A color video graphics array monitor displayed a real-time tracing of exhaled volume vs time, which was viewed by the subject and the technician. Numeric results relevant to criteria 2 to 4 were displayed after each maneuver, and results were relevant to criterion 1 after all but the first maneuver. Data from all maneuvers were stored on the hard disk (and copied to a floppy disk) for subsequent analysis.

A QA officer (W.S.L.) made unannounced visits to observe field testing and reviewed the collected data to look for consistent differences between technicians in the incidence of unsatisfactory test performances. In years 1 to 3, evaluations were performed *ad hoc*, based on the numeric data generated by the data acquisition program. Beginning in year 4, the central QA review and the reporting software developed for the Lung Health Study (LHS)[6] were implemented, in parallel to the usual data acquisition. Thereafter, the QA officer used the review software to inspect all flow-volume tracings from subjects with questionable performance, overriding the selection of the spirometer software algorithm of the "best" maneuver on rare occasions. The software then stored performance data for each subject, calculated the average flow and volume quality grades for all subjects tested by an individual technician, and generated monthly written reports for technicians and investigators.

For this review, we analyzed data obtained via the QA software in years 4, 5, and 6, which covered 4,143 subjects, including the grade 4 cohort recruited in year 4 (typical age in year 4, 9 years [with no prior lung function testing experience]), the original grade 4 cohort (typical age in year 4, 12 years [with 3 years of prior experience]), and the original grade 7 cohort (typical age in year 4, 15 years [with 3 years of prior experience]). Thus, the review included both naive subjects near age 9 years and older, experienced subjects up to age 18 years.

Of all subjects, 49% were boys, 34% belonged to ethnic minorities (mostly Mexican Americans, African Americans, and Asian Americans), 14% had a history of physician-diagnosed asthma, and an additional 19% had a history of wheeze without a diagnosis of asthma. These subjects performed a total of 9,528 test sessions. Because of personnel turnover, this analysis included 11 different technicians, both new and experienced, with a wide range of educational and occupational backgrounds. An additional three short-term substitute technicians, who together performed < 1.8% of all tests, had spirometry quality similar to the regular technicians. However, their results were excluded from the analysis to limit the number of regression variables required (see below). An additional 1.7% of all tests were included in descriptive statistics but not in regression analyses, because data were missing for one or more predictor variables.

To assess the overall quality of performance by our subjects and technicians, descriptive statistics were calculated for the maneuver acceptability variables BEV, time to peak expiratory flow (PEFT), EOTV, and FET for the single best maneuver from each test session, as determined by the QA software, and for differences between the best and second-best FVC (dFVC), FEV_1 ($dFEV_1$), and PEF (dPEF) values within a spirometry test session. These reproducibility variables, which were not stored by the QA software, were extracted from a separate Children's Health Study database that covered years 1 to 5, and the best maneuvers were selected separately for each spirometry variable. Thus, differences were analyzed over 2 years (years 4 and 5), rather than over 3 years, and were not necessarily based on a single composite best maneuver, as was the case for the maneuver-acceptability variables.

To identify significant influences on performance, multiple regression analyses were performed on each performance-quality variable. The initial regressions included continuous independent variables for year, age, height, and body mass index (kilograms/meters squared; BMI), as well as dichotomous independent variables (false, 0; true, 1) for male gender, ethnic minority status, naive subject (first testing experience), and home testing. (In preliminary analyses, the effect of new vs experienced technicians was tested and found to be unimportant, so that variable was excluded.) Additional dichotomous independent variables represented each individual technician, using one technician with near-average performance as a reference. From the initial result, independent variables with near-significant ($p < 0.15$) or significant ($p < 0.05$) effects were identified and entered into a second regression, which yielded the final model.

Because data distributions of QA variables tended to be skewed, with longer tails on the side of poorer performance, alternative regression analyses were performed after log transformations to obtain more nearly normal distributions, or with the most "outlying" poor performers excluded (described below). For the variables derived from the best single blow (*ie*, maneuver-acceptability checks), conclusions from the original and alternative regression analyses were not substantially different; accordingly, the original untransformed variables are reported. However, for the reproducibility variables, the original and the alternative analyses differed, indicating that original analyses were heavily influenced by a few outliers. Accordingly, for the difference variables, we report the results from analyses that exclude the poorest performances, which were defined as > 10% dFVC or $dFEV_1$, or > 20% dPEF.

Downloaded From: http://journal.publications.chestnet.org/pdfaccess.ashx?url=/data/journals/chest/21952/ on 01/05/2017

Analyses were performed using statistical software (BMDP-DYNAMIC; SPSS Inc; Chicago, IL; or SAS, version 6; SAS Institute; Cary, NC).

Results

Table 1 gives descriptive statistics for each performance variable, based on 9,355 test sessions over 3 years (5,561 test sessions over the earlier 2 years for best vs second-best differences). The percentages of test sessions failing to meet each ATS and LHS (adult-based) maneuver-acceptability criterion and each reproducibility criterion were the following: EOTV, 2%; FET, 5%; BEV, 7%; dFVC, 3%; $dFEV_1$, 7%; and dPEF, 10%.

Table 2 summarizes regression results for the four performance variables from the best single maneuver, based on 9,199 test sessions for which complete predictor data were available. Blanks indicate effects that were not near significance ($p > 0.15$) in initial analyses and were excluded from final analyses. Maturity and experience clearly improved performance. The naive-subject effect (*ie*, the comparison of year 4 tests for new grade 4 subjects against all other tests) was significant in the unfavorable direction for all four variables. PEFT and EOTV improved significantly in older subjects relative to younger ones, regardless of the year of testing; PEFT also showed overall improvement from year to year. Taller subjects showed significantly lower BEVs, faster PEFTs, and longer FETs. More obese subjects, as judged by BMI, had slower PEFTs and larger EOTVs despite having longer FETs. Subjects with histories of asthma or wheeze were significantly better performers than other subjects, showing lower BEVs and PEFTs as well as longer FETs.

Boys showed significantly better performance than girls with respect to BEV, PEFT, and FET. Supplementary analyses (not tabulated) showed that this gender difference was not influenced significantly by the technician's gender. That is, boys performed about equally well for female and for male technicians, and the same was true of girls. Significant differences among individual technicians were uncommon, except for FET. No technician significantly exceeded the "reference" performance level for all four test variables, although technician B significantly exceeded it for all except EOTV. Home testing was associated with significantly shorter FET compared to testing at schools but otherwise had no important influence on performance.

It was more difficult for the children to meet the PEFT criteria (used by the LHS) than the BEV criteria (recommended by the ATS). The girls' PEFTs were > 120 ms in 15% of tests, while the boys' PEFTs were > 90 ms in 32% of tests. Of those subjects with an "unacceptably" high PEFT, 17% met the BEV criteria.

Ethnic minority subjects took longer to reach peak flow and had larger BEVs than non-Hispanic white (*ie*, ethnic majority) subjects, but their EOTV and FET values were essentially the same. Later analyses of specific ethnic categories (not tabulated) explored the BEV differences in more detail. Ethnic majority subjects showed significantly lower BEVs if tested by ethnic majority technicians than if tested by ethnic minority technicians. Subjects who classified themselves as Hispanic or of mixed ethnicity (*ie*, mostly of Mexican or Central American ancestry, including some native Spanish speakers) showed significantly lower BEV values if tested by bilingual Hispanic technicians than if tested by other technicians. Too few data were available to test for similar effects in other ethnic minority groups.

Table 3 summarizes regression results for the reproducibility measures (*ie*, dPEF, $dFEV_1$, and dFVC), based on 5,457 test sessions in the earlier 2 years with complete predictor data (excluding outli-

Table 1—*Distributions of Performance Variables**

	Percentile						
Variable	5th	10th	25th	50th	75th	90th	95th
BEV, %	1.3	1.9	2.7	3.4	4.0	4.8	5.7
PEFT, ms	62	66	75	87	101	132	157
EOTV, mL	−32	−24	−12	−1	17	37	56
FET, s	4.0	4.6	5.6	6.6	7.6	8.6	9.4
dFVC, %	0.0	0.1	0.4	1.0	2.0	3.2	4.2
dFVC, mL	0	4	11	31	61	98	132
$dFEV_1$, %	0.0	0.2	0.5	1.2	2.5	4.3	6.4
$dFEV_1$, mL	0	7	11	32	63	117	171
dPEF, %	0.3	0.5	1.3	3.0	5.8	10.0	14.2
dPEF, mL/s	15	28	77	177	355	610	820

*N = 9,355 test sessions for the four maneuver acceptability variables, and N = 5,561 for the reproducibility variables.

Downloaded From: http://journal.publications.chestnet.org/pdfaccess.ashx?url=/data/journals/chest/21952/ on 01/05/2017

Table 2—*Significant ($p < 0.05$) Predictors of Spirometric Performance Quality by Four Separate Criteria Applied to the Best Single Maneuver

Predictor	BEV, %	PEFT, ms	EOTV, mL	FET, s
Year		−2.02/yr		
Age		−0.55/yr	−3/yr	
Height	−0.015/cm	−0.12/cm		0.004/cm
BMI		0.5/U	1/U	0.08/U
Male gender	−0.42	−8.7		0.48
Asthma/wheeze	−0.16	−2.2		0.18
Ethnic minority	0.25	5.3		
Naive subject	0.27	6.5	38	−0.28
Home test	−0.26			−0.37
Technician A			29	
Technician B	−0.40	−7.4		0.45
Technician C	0.09			−0.26
Technician D			30	
Technician E				0.93
Technician F	−0.16			−0.11
Technician G				1.50
Technician H			12	0.47
Technician J	−0.21	−4.7		−0.24
Technician K				+0.82
Technician L	0.44			

*N = 9,199 test sessions.

ers). The initial models included all predictors listed in Table 2; only those that were significant for at least one reproducibility measure are listed in Table 3. Ethnic majority subjects showed better reproducibility than minority subjects for all three function measures. Taller subjects showed more reproducible FVC and PEF values, while more obese subjects (*ie*, those with a higher BMIs) showed less reproducible PEF values. Few significant differences were seen in the ability of individual technicians to obtain reproducible PEF, FEV_1, and FVC values.

All the aforementioned statistically significant effects were small in magnitude and explained little of the variance in performance. Only 20% of the variance in FET, and < 10% of the variance in any other QA measure, was explained by the combined influence of all predictors used in regression analyses. Greater predictability of FET resulted from individual differences among technicians; however, those differences appeared relatively unimportant to FVC accuracy, in light of the generally nonsignificant differences in EOTV. In an analysis ignoring other factors, maneuver duration (*ie*, FET variance) explained < 4% of EOTV variance. On the other hand, PEFT variance explained about 50% of BEV variance.

Table 3—*Significant Independent Predictors of Reproducibility for FVC, FEV_1, and PEF

Predictor	dFVC, %	$dFEV_1$, %	dPEF, %
Height	−0.0053/cm		−0.020/cm
BMI			0.031/U
Male gender			−0.248
Ethnic minority	0.075	0.168	0.389
Technician A	−0.175		
Technician D			0.683
Technician H	−0.163		
Technician J	0.206		

*Candidate predictors included all those listed in Table 2. N = 5,457 test sessions (after excluding outliers).

Discussion

In general, the spirometry quality of the children and adolescents in our study compared favorably with that of adults studied elsewhere.[7–13] The thresholds specified by the ATS[1] and LHS[6] spirometry acceptability and reproducibility criteria were set so that about 5% of adults failed to meet each criterion when tested by an experienced technician using a diagnostic-quality spirometry system. Our results show that children aged ≥ 9 years also can meet each of the ATS criteria about 95% of the time.

Automated maneuver quality checks with messages displayed immediately after each maneuver enable the technician (or patient) to perform additional maneuvers, increasing their ability to meet the test session quality goals.

The spirometry maneuver may be divided into the following three steps (or phases), each of which requires a different type of effort: (1) "take a deep breath" (maximal inhalation); (2) "blast out your air" (maximal exhalation effort); and (3) "keep blowing until all your air is gone" (prolonged exhalation). Poor effort may occur during any (or all) of these steps and is usually due to suboptimal interaction between the technician and the subject. A submaximal inhalation falsely reduces the PEF, FEV_1, and FVC values. A submaximal blast falsely reduces the PEF values, variably affects the FEV_1 values, and may increase the FVC values. A premature termination of the exhalation falsely reduces only the FVC values, provided that termination occurs after the first second. Objective QA measurements are designed to detect all these faults and, thereby, to identify any poorly performed maneuver or test session that could result in false-positive or false-negative diagnoses in the clinical setting or in increased measurement noise/bias in epidemiologic and intervention studies.

Poor inhalation effort is common but is not objectively evident in any single spirometric record, being detectable only by invasive physiologic measurements or possibly by subjective visual observation of the subject's performance. Thus, for practical purposes, poor inhalation effort can be detected only in

Downloaded From: http://journal.publications.chestnet.org/pdfaccess.ashx?url=/data/journals/chest/21952/ on 01/05/2017

terms of poorly reproducible FVC and FEV_1 values across multiple maneuvers. Submaximal blast and premature termination can, however, be identified objectively from the recording of any single blow. The following sections summarize findings in this and earlier studies, and our recommendations based on these findings, concerning each type of fault.

Criteria to Detect Poor Reproducibility

We found that a history of asthma or wheeze predisposed our subjects to better performance, in contrast to several adult studies that indicate that poor performance is more common in individuals with clinical evidence of impaired respiratory health.[7,10,13] Ng'ang'a and coworkers[12] noted that young men with bronchial hyperresponsiveness and young women who were cigarette smokers were more likely than others to fail to meet the $dFEV_1$ criterion of < 0.1 L. About 12% of their subjects failed to meet that reproducibility criterion. In a large α_1-antitrypsin disease registry at 37 sites, Stoller and coworkers[13] noted that only 2% of the participants failed to meet the $dFEV_1$ criterion. Spirometry of subjects who failed to meet the criterion was more likely to occur at those sites using older water-sealed spirometers without personal computer displays of flow-volume curves, but the failure to meet the criterion was not associated with asthma, gender, or age. Their patients with severe airways obstruction were much less likely to fail to meet the $dFEV_1$ criterion than those with relatively normal lung function.

Most of our study subjects with a history of wheezing or asthma had only mild, if any, manifestations of disease at the times of testing. We do not know how many also performed spirometry or peak flow maneuvers during visits to their private physicians (which would potentially improve their test quality). Whether similarly high levels of performance can be obtained in the clinical testing of children seen for respiratory diseases is less certain. In clinical populations, illness and higher anxiety levels may impair performance more frequently. On the other hand, fewer distractions and the ability to take more time might offset these factors.

Based on our 95th percentile results (Table 1), we recommend that the 1994 ATS reproducibility criterion of striving for a < 0.20-L difference between the highest and second-highest values ($dFEV_1$ and dFVC) apply when testing children (Table 4). About 95% of our subjects also met the more stringent 1987 ATS criteria of a < 5% difference. We also recommend that when PEF is being measured using a spirometer, the reproducibility criteria of dPEF of < 1.0 L/s and < 15% should be applied when testing children and adolescents.

Table 4—*Recommended Spirometry Test Session Acceptability and Reproducibility Goals When Testing Children Ages 9 to 18 Years*

QA Variables	Goals
BEV	< 5% of the FVC
PEFT	< 160 ms
EOTV	< 60 mL*
FET	> 6 s
dFVC	< 200 mL and < 5%
$dFEV_1$	< 200 mL and < 5%
dPEF	< 1.0 L/s and < 15%

*This criterion may not apply to flow-sensing spirometers or to children with airways obstruction.

Criteria to Detect a Slow Start

The second phase of the spirometry maneuver is to blast out the air as quickly as possible, thereby achieving a "sharp" (high) peak flow during the first 10th of a second and a high average flow during the first second of the maneuver (*ie*, FEV_1). A hesitating start creates a high BEV, and the FEV_1 may then be underestimated. A long PEFT indicates a relatively slow start or the lack of a maximal effort to blast out the air. It is important to use both the BEV and PEFT criteria since a patient may have a short PEFT and high PEF following a hesitating start (*ie*, a large BEV), or the patient may have an acceptably low BEV followed by a sigh (*ie*, large PEFT, low PEF, and falsely low FEV_1 values).

The criteria for an acceptably short PEFT developed by the LHS from smoking adults (women, < 90 ms; men, < 120 ms) appear too strict when applied to the young people in our study. Miller and coworkers[14] measured the PEF 10 to 90% rise time in adult patients using a pneumotach (flow sensor)-based spirometer and found that the values from 95% of the patients were < 140 ms. Their patients with airway obstruction had considerably shorter rise times when compared to those with relatively normal lung function. For a given maneuver, the measured PEFT will be slightly larger than the 10 to 90% rise time. For children, we recommend using a PEFT threshold of > 160 ms to detect a "slow start," since 95% of our subjects produced a PEFT less than this value. More so than the other QA criteria, the measured PEFT may vary from system to system. When tested in Salt Lake City using a waveform generator, our spirometers overestimated PEF values at higher PEF values, suggesting that the systems were underdamped. Therefore, it is likely that

Downloaded From: http://journal.publications.chestnet.org/pdfaccess.ashx?url=/data/journals/chest/21952/ on 01/05/2017

the relatively longer PEFTs obtained in our study were due to less phase II "blast" effort than to an underdamped measuring system. The ATS did not mention PEFT acceptability criteria for diagnostic spirometry in 1994, but the PEFT, dPEF, and $dFEV_1$ were all independent predictors of month-to-month variability in FEV_1 in the LHS.[2]

The ATS recommendations stated that: "Although there may be some benefit from using PEF reproducibility to improve subject effort, no specific [PEF] reproducibility criterion is recommended at this time."[1] Coates and coworkers[8] found that in children with asthma or cystic fibrosis who were tested by hospital-based pulmonary function laboratory technicians, that variation in FEV_1 (*ie*, $dFEV_1$) was much more closely associated with the variation in the FVC due to variation in the depth of inhalation preceding the FVC maneuver than with the variation in peak flow (*ie*, dPEF). Banks and coworkers[15] used experienced technicians to study adult employees with two different spirometry systems and found that the newer system, which had automated maneuver quality checks and messages (*ie*, dPEF, EOTV, BEV, and an index similar to PEFT), resulted in higher FVC values (mean increase, 0.27 L) but no significant change in FEV_1 values.

The 1994 ATS recommendations stated: "Computer-based systems that provide feedback to the technician when the above conditions [acceptability and reproducibility criteria] are not met are desirable." As hand-held electronic spirometers increase in popularity for ambulatory lung function measurements for clinical trials of asthma therapy that use PEF and FEV_1 values as primary outcome measures of asthma control,[16] automated acceptability and reproducibility maneuver quality checks such as PEFT, dPEF, and BEV, which detect poor-quality PEF and FEV_1 results, will become important. In the ambulatory setting, the study subjects test themselves without a technician or a display of flow-volume curves to recognize unacceptable maneuvers.

Criteria to Detect a Short Effort

End-of-test maneuver acceptability criteria are designed to detect maneuvers that "quit too soon" resulting in an underestimation of the true FVC. The 1987 ATS recommendations[17] required FET > 6 s, and an "obvious plateau" in the volume-time curve, defined as a < 40-mL volume change during the final 2 s of the maneuver (*ie*, EOTV). The European Respiratory Society decided not to recommend EOTV criteria,[18] apparently since patients with moderate-to-severe airways obstruction can forcefully exhale for > 20 s and still not reach a plateau, while healthy children often reach a plateau in < 3 s. Desmond and coworkers[19] found that > 80% of the children in their study (age range, 5 to 18 years) failed to meet the 1987 ATS EOTV criteria (based on adults), while only 9% failed the ATS $dFEV_1$ or dFVC criteria. They proposed a revised EOTV criterion using exponential curve fitting of the volume-time data from each maneuver. School-aged children failed to meet the new criterion only 4% of the time, but preschool children (age, < 7 years) failed to meet it about 37% of the time. About 60% of the adult COPD patients in one large study failed to meet the 1987 ATS EOTV criteria.[13] About 95% of our subjects exhaled < 56 mL during the last 2 s of the maneuver. Children performing spirometry for the first time in our study had a mean EOTV that was 38 mL higher than those who had performed spirometry before. Three of the technicians (technicians A, D, and H) were less likely to obtain plateaus than the other technicians, but, surprisingly, the mean expiratory times they obtained from the children they tested were not significantly shorter.

It is easier to attain a low EOTV when using volume spirometers such as the dry-rolling seal that we used (when compared to flow-sensing spirometers), since the warm, exhaled air contracts when it cools quickly inside the spirometer, making it appear that the subject has reached a plateau even though very slow exhalation continues toward the end of the maneuver. Note that the EOTV was negative for half of our subjects (Table 1), indicating contraction of the exhaled air (not slow inhalation or a leak).

The ATS reduced the importance of the EOTV criteria in 1994, stating that "a plateau should be observed, as defined by no change in volume for at least one second OR a reasonable expiratory time. . . usually 6 seconds." About 35% of our subjects had an exhalation time of < 6 s during their best maneuver, but only 5% had an FET of < 4 s. We propose that, when testing population samples of children, the message "premature termination of effort" be displayed when the FET is < 4 s or when the EOTV is > 60 mL, but that the EOTV criterion should not be applied when testing children known to have asthma or airways obstruction. Luckily, underestimation of the FVC is not clinically important when following patients or study subjects with asthma. The forced exhalation may stop after only 1 s when only the measurement of the PEF or FEV_1 is sought.

Our results also suggest that ethnic and linguistic matching between subject and technician increases the probability of satisfactory performance but that gender matching has little influence. By some criteria, boys performed significantly better than girls, and non-Hispanic whites performed better than ethnic minority subjects. However, the differences were small, and almost all of the children from each

Downloaded From: http://journal.publications.chestnet.org/pdfaccess.ashx?url=/data/journals/chest/21952/ on 01/05/2017

group provided satisfactory data. Hankinson and Bang[9] also noted that ethnicity was an independent factor associated with $dFEV_1$ in the third National Health and Nutrition Survey population-based study.

Summary

We found that most schoolchildren who are ≥ 9 years can perform forced expiratory spirometric maneuvers well enough to meet currently established adult-based maneuver acceptability criteria for BEV, EOTV, and FET, as well as within-test session maneuver reproducibility for FEV_1, FVC, and PEF. Factors related to age, size, gender, ethnicity, and subject-technician affinity can influence performance, but their overall effect is small with well-trained technicians who have experience. Our findings suggest the need for criteria to be added to those currently recommended by the ATS to ensure adequately fast starts and reproducible PEF values that are applicable to adults as well as to children. For 9- to 18-year-old children, we recommend the spirometry test session goals (criteria) listed in Table 4.

ACKNOWLEDGMENTS: We acknowledge the dedication and skills of the Childrens Health Study pulmonary function technicians, and the support of the school district offices, administrators, and teachers from the participating schools. We especially thank the students and parents involved in the study who allowed us to work with them and to learn about them.

References

1 American Thoracic Society. Standardization of spirometry: 1994 update. Am J Respir Crit Care Med 1995; 152:1107–1136
2 Peters JM, Avol EL, Navidi W, et al. A study of twelve southern California communities with differing levels and types of air pollution: I. Prevalence of respiratory morbidity. Am J Respir Crit Care Med 1999; 159:760–767
3 Peters JM, Avol EL, Gauderman WJ, et al. A study of twelve southern California communities with differing levels and types of air pollution: II. Effects on pulmonary function. Am J Respir Crit Care Med 1999; 159:768–775
4 Dockery DW, Speizer FE, Stram DO, et al. Effects of inhalable particles on respiratory health of children. Am Rev Respir Dis 1989; 139:587–594
5 Linn WS, Solomon JC, Gong H, et al. Standardization of multiple spirometers at widely separated times and places. Am J Respir Crit Care Med 1996; 153:1309–1313
6 Enright PL, Johnson LR, Connett JE, et al. Spirometry in the Lung Health Study: methods and quality control. Am Rev Respir Dis 1991; 143:1215–1223
7 Becklake MR. Epidemiology of spirometry test failure. Br J Ind Med 1990; 47:73–74
8 Coates AL, Desmond KJ, Demizio D, et al. Sources of variation in FEV_1. Am J Respir Crit Care Med 1994; 149: 439–443
9 Hankinson JL, Bang KM. Acceptability and reproducibility criteria of the ATS for spirometry as observed in a sample of the general population. Am Rev Respir Dis 1991; 143:516–521
10 Humerfelt S, Eide GE, Kvale G, et al. Predictors of spirometric test failure: a comparison of the 1983 and 1993 acceptability criteria from the ECCS. Occup Environ Med 1995; 52:547–553
11 Kunzli N, Ackermann-Liebrich U, Keller R, et al. Variability of FVC and FEV_1 due to technician, team, device, and subject in an eight center study. Eur Respir J 1995; 8:371–376
12 Ng'ang'a LW, Ernst P, Jaakkola MS, et al. Spirometric lung function: distribution and determinants of test failure in a young adult population. Am Rev Respir Dis 1992; 145:48–52
13 Stoller J, Buist A, Burrows B, et al. Quality control of spirometry testing in the registry for patients with severe alpha-1-antitrypsin deficiency. Chest 1997; 111:899–909
14 Miller MR, Pedersen OF, Quanjer PH. The rise and dwell time for peak expiratory flow in patients with and without airflow limitation. Am J Respir Crit Care Med 1998; 158: 23–27
15 Banks DE, Wang ML, McCabe L, et al. Improvement in lung function measurements using a flow spirometer that emphasizes computer assessment of test quality. J Occup Environ Med 1996; 279–283
16 Enright PL, Lebowitz MD, Cockroft DW. Asthma Outcome Measures Workshop: pulmonary function tests. Am Rev Respir Dis 1994; 149(suppl):S9–S18
17 American Thoracic Society. Standardization of spirometry: 1987 update. Am Rev Respir Dis 1987; 136:1285–1293
18 Quanjer PH, Tammeling GJ, Cotes JE, et al. Lung volumes and forced ventilatory flows: report of working party, standardization of lung function tests. Eur Respir J 1993; 6(suppl):5–40
19 Desmond KJ, Allen PD, Demizion DL, et al. Redefining end of test (EOT) criteria for pulmonary function testing in children. Am J Respir Crit Care Med 1997; 156:542–545

Downloaded From: http://journal.publications.chestnet.org/pdfaccess.ashx?url=/data/journals/chest/21952/ on 01/05/2017